

## Упрощенный метод анализа, контроля и прогноза высокого загрязнения воды отдельными компонентами

**\*О.М. Розенталь<sup>1</sup>, Л.Н. Александровская<sup>2</sup>, А.В. Кириллин<sup>2</sup>**

<sup>1</sup>Институт водных проблем РАН, Российская Федерация, 119991, г. Москва, ул. Губкина, д. 3

<sup>2</sup>Московский авиационный институт,  
Российская Федерация, 125993, г. Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4.

\*Адрес для переписки: Розенталь Олег Моисеевич, E-mail: orosental@rambler.ru

Поступила в редакцию 3 декабря 2016 г., после доработки – 11 мая 2017 г.

Оценка риска и прогноз повышенного загрязнения природных вод особенно важны в промышленных регионах, где нестабильность контролируемых показателей наиболее высока. Поскольку риск является вероятностной характеристикой степени загрязнения воды отдельными компонентами, для решения этой задачи необходимо привлечение статистических методов. В статье предложена практическая методика такой оценки с использованием непараметрических, основанных на теории порядковых статистик, и более точных параметрических методов статистического анализа данных гидромониторинга, требующих аналитического описания функции распределения концентрации загрязняющих веществ. Результаты использования методики апробированы для уральского региона, для чего обосновано и показано, что аналитическое описание «хвостовых» частей функций распределения вероятностей при больших значениях исследуемых показателей качества воды распределением Парето позволяет по выборкам ограниченного объема определять соответствующие показатели в ненаблюдаемых частях распределений, т.е. давать прогнозные оценки. Отмечено, что риск высокого загрязнения воды существует, даже если ситуация по данным гидромониторинга относительно благополучна. В связи с этим предложена методика оценки доверительного интервала, покрывающего значения контролируемого показателя с заданной доверительной вероятностью, в том числе наиболее важной его верхней доверительной границы. На практических демонстрационных примерах показана эффективность разработанной методики. Особая значимость применения предложенных подходов к прогнозной задаче обусловлена возможностью оценивания малых рисков появления экстремальных единичных значений исследуемых случайных величин.

**Ключевые слова:** вероятность высокого загрязнения воды, прогноз, «хвост» функции распределения вероятностей, ранжирование наибольших значений концентрации, верхняя граница доверительного интервала, теория порядковых статистик

For citation: *Analitika i kontrol'* [Analytics and Control], 2017, vol. 21, no. 2, pp. 153-160

DOI: 10.15826/analitika.2017.21.2.008

## Simplified method of analysis, control and forecast of high water pollution by separate components

**\*O.M. Rosental'<sup>1</sup>, L.N. Aleksandrovskaya<sup>2</sup>, A.V. Kirillin<sup>2</sup>**

<sup>1</sup>Institute of Water Problems of RAS, Gubkin ul., 3, Moscow, 119991, Russian Federation

<sup>2</sup>Moscow aviation Institute, Volokolamskoe shosse, 4, Moscow, 125993, Russian Federation

\*Corresponding author: Oleg M. Rosental', E-mail: orosental@rambler.ru

Submitted 03 December 2016, received in revised form – 11 may 2017

Assessing the risks and predicting the increased pollution of natural water bodies is especially important in the industrial regions where the instability of the monitored indicators is most significant. This article proposes methods of this activity with the use of non-parametric and more accurate parametric methods of hydro-monitoring data analysis. The results of using these methods for the Ural region have been obtained, and it has been shown that the analytical description of the "tail" parts of the probability distribution in the

studied water quality indicators with the use of the Pareto distribution allows to identify the appropriate indicators for the limited volume samples and to give the projected estimates in the unobserved parts of the distributions. It has been noted that the risk of significant water pollution exists even if the situation is relatively favorable according to the hydro-monitoring. In this regard, a methodology of estimating the upper limit of the confidence interval covering the value of the monitored index with the acceptable reliability has been proposed. Practical examples show the effectiveness of this method. The special significance of the application of the proposed approaches to the predicting aspect is due to the possibility of assessing the small risks of the appearance of extreme unit values of the investigated random variables.

**Keywords:** probability of considerable water contamination, prediction, the «tail» of the probability distribution function, ranking the highest concentrations, upper fiducial limit, the theory of order statistics.

## Введение

Оценка риска высокого загрязнения (**ВЗ**) воды водных объектов, ее экстремально высокого загрязнения (**ЭВЗ**) и прогнозирование этих событий исключительно важны для устойчивого водопользования. Соответствующие руководящие документы [1, 2], критерии оценки опасности токсического загрязнения [3] и классификации категорий аварийной ситуации по степени опасности ожидаемых последствий [4] только выиграют при их сочетании с необходимыми статистическими методами исследования на основе вероятностных моделей. В частности, для корректной оценки ВЗ и ЭВЗ полезно установление доверительного интервала, прежде всего, его верхней границы. В свою очередь с этой целью желательно знание закона распределения вероятностей искомых характеристик, особенно его «хвостовой части».

### 1. Анализ функции распределения вероятностей концентрации загрязняющих веществ

Концентрация загрязняющих веществ воды в рамках концепции приемлемого риска считается случайной величиной, для наиболее полного описания которой используется функция (закон) распределения вероятностей соответствующих компонент загрязнения.

В разных случаях закон распределения вероятности концентрации загрязняющих веществ представляется степенной или экспоненциальной функцией [5-7]. Наиболее распространенным экспоненциальным законом является нормальный, который реализуется, если влияющие на качество воды факторы естественной и антропогенной природы многочисленны и независимы, их значимость – соразмерна, а характер воздействия – аддитивный. На практике эти условия выполняются не всегда из-за причин неслучайной (нештатной) природы, «залповых» сбросов загрязняющих воду веществ, смывов, снеготаяния, других нежелательных природных и техногенных явлений. В результате происходит появление больших вероятностей в правой ветви эмпирического распределения, построенного по экспериментальным данным, и появление в его «хвостовой» части скачков значений вероятности контролируемого показателя. Такие функции распределения вероятностей носят на-

звание распределений с «тяжелыми хвостами» [8, 9]. Эти функции распределения не подчиняются нормальному закону, что хорошо известно в гидрологии [9].

Имеются многочисленные попытки описать законы распределения с так называемыми «тяжелыми хвостами», например гамма-распределением или распределением Вейбулла [10, 11]. Однако используемый при этом математический аппарат достаточно сложен, что затрудняет доведение полученных результатов до практических методик. Кроме того, механизм образования экстремальных событий отличается от стабильного протекания гидрохимических процессов, что ставит под сомнение возможность аппроксимации эмпирической функции распределения каким-либо одним «чистым» законом. В статье предлагается оригинальный подход, при котором «подгонка» математических статистических моделей осуществляется только в интервале, где располагаются интересующие нас значения концентрации с дальнейшей экстраполяцией полученных результатов на ненаблюдаемые хвосты распределений. Для аппроксимации правых «хвостов» распределения могут быть использованы различные законы, ограниченные слева. Простейшим из таких законов является распределение Парето, описывающее поведение случайных величин, больших некоторого фиксированного значения [9, 12].

Заметим, что проведенные ниже исследования по возможности использования распределения Парето для описания «хвостов» распределений и согласование его с усеченной справа эмпирической частью распределения применимы для широкого класса возможных видов загрязняющих веществ, однако приведенные ниже примеры, представляющие собой по существу прикладные методики, носят демонстрационный характер.

**Пример 1<sup>1</sup>.** Почти повсеместный избыток железа в речной воде промышленных регионов создает проблемы для ее промышленного и бытового использования, а в случаях ЭВЗ часто угнетает развитие водных экосистем. Поэтому представляет интерес продемонстрировать возможность описания экспериментальных данных нормальным рас-

1 Для целей данного и дальнейших примеров использованы массивы экспериментальных данных Росгидромета, предоставленные авторам ФГБУ ГХИ в Ростове-на-Дону.

пределением и распределением Парето для этого вещества, используя результаты ежедневного контроля его содержания в воде р. Исеть (приток Тобола) на контрольном створе 9.3 км ниже г. Каменск-Уральского в 1999–2001 гг. в количестве  $n_0 = 824$  [7].

**Решение.** Исследование применимости нормального распределения.

Результаты обработки экспериментальных данных показывают, что оценки математического ожидания  $\hat{m} = 0,48$  мг/дм<sup>3</sup> и среднеквадратического отклонения  $\hat{\sigma} = 0,48$  мг/дм<sup>3</sup> соизмеримы. Следовательно, значительная часть левой ветви эмпирического распределения при условии его нормальности переходила бы в область физически невозможных отрицательных значений контролируемой концентрации. При этом вместо свойственных нормальному распределению значений показателей асимметрии  $\sqrt{\hat{\beta}}_1 = 0$  и эксцесса  $\hat{\beta}_2 = 3$  имеем 2.576 и 12.244, соответственно. Следовательно, экспериментальное распределение, в отличие от нормального, не симметричное и более «островершинное». Положительная асимметрия указывает на растянутую правую ветвь графика функции распределения вероятностей, т.е. «тяжелый» правый «хвост», так что, по крайней мере, в этой части распределения использование нормального закона неприемлемо.

Поскольку данная работа посвящена проблеме прогнозирования высокого загрязнения вод, ниже рассмотрена применимость степенного распределения для описания указанной «хвостовой части», а именно распределения Парето  $F_{\text{Парето}}(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha$ , где  $x$  – контролируемый показатель,  $x_0 \leq x$  – граничное значение,  $\alpha$  – параметр распределения.

В данном случае, следуя [12], из имеющейся выборки  $n_0$  выделены, например,  $n = 20$  наибольших значений концентрации  $x_{(i)}$ , где  $i = 1, 2, \dots, 20$  (строки 2 и 5 в табл. 1) и рассчитаны соответствующие экспериментальные вероятности  $F_i = \frac{i}{n+1}$  [13] – (строки 3 и 6).

Наиболее важен вопрос о сопоставимости максимального экспериментального значения  $x_{(20)}$  с соответствующим расчетным  $\hat{x}_{(20)}$ . Для оценки последней величины необходимо определить значение параметра  $\alpha$ . С этой целью по данным табл. 1 рассчитаны:

– среднее значение измеренной концентрации

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{(i)} = 2.29 \text{ мг/дм}^3;$$

– дисперсия  $S^2 = \frac{1}{n} \sum_{i=1}^n (x_{(i)} - \bar{x})^2 = 0.19 \text{ мг}^2/(\text{дм}^3)^2$ ;

– оценка коэффициента вариации  $\hat{v} = \frac{S}{\bar{x}} = 0.04$ .

Отсюда оценка искомого параметра  $\hat{\alpha}$ , полученная методом моментов, путем приравнивания теоретических моментов распределения и их оценок,

$$\text{следующая } \hat{\alpha} = 1 + \sqrt{1 + \left(\frac{1}{\hat{v}}\right)^2} = 6.32.$$

При подстановке этих данных в распределение Парето (для этого удобна его форма в виде  $\ln(1 - F_{(20)}) = \alpha \ln(x_0/\hat{x}_{(20)})$ ), и принятом в результате округления  $x_{(1)}$  в меньшую сторону  $x_0 = 1.8$ , получено расчетное значение  $(x_0/\hat{x}_{(20)}) = 0.618$ , откуда,  $\hat{x}_{(20)} \approx 2.91$ , что меньше экспериментального значения  $x_{(20)} = 3.12$ , т.е. имеет место недооценка уровня концентрации  $\Delta = \frac{3.12 - 2.91}{3.12} \cdot 100\% = 6.73\%$  мень-

шая погрешности измерения [6], поэтому точность аппроксимации удовлетворительная.

На практике собрать 20 редких событий затруднительно. На основе аналогичных расчетов для 10 ранжированных наибольших значений концентрации, получено:  $\bar{x} = 2.63$  мг/дм<sup>3</sup>;  $S^2 = 0.13 \text{ мг}^2/(\text{дм}^3)^2$ ;  $\hat{v} = 0.02$ ;  $\hat{\alpha} = 8.26$ ;  $F_{(10)} = 10/11$ ;  $x_0 = 2.12$  мг/дм<sup>3</sup>. И тогда  $x_{(10)} \approx 2.84$  мг/дм<sup>3</sup>. Следовательно, результат достаточно устойчив.

**Примечание к примеру.** Возможно повышение точности аппроксимации путем расчета параметра  $\alpha$  методом наименьших квадратов, нивелирующим факт низкой точности оценок  $\bar{x}$  и  $S^2$ , а следовательно, и  $\hat{\alpha}$  по малому объему выборки

**Таблица 1**

Наибольшие значения концентрации железа (мг/дм<sup>3</sup>) в р. Исеть (экспериментальные данные) и их соответствующие экспериментальные вероятности [7]

**Table 1**

Highest values of the concentration of iron (mg/dm<sup>3</sup>) in the Iset' River (experimental data) and their corresponding experimental probabilities [7]

$I$	1	2	3	4	5	6	7	8	9	10
$x_{(i)}$	1.81	1.82	1.83	1.89	1.92	1.95	2.01	2.06	2.07	2.09
$F_i$	1/21	2/21	3/21	4/21	5/21	6/21	7/21	8/21	9/21	10/21
$i$	11	12	13	14	15	16	17	18	19	20
$x_{(i)}$	2.13	2.15	2.2	2.34	2.75	2.76	2.85	2.94	3.08	3.12
$F_i$	11/21	12/21	13/21	14/21	15/21	16/21	17/21	18/21	19/21	20/21

( $n = 20, n = 10$ ). Условие поиска минимума следующее:  $\min \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i)^2$ . Здесь в соответствии с формулой Парето

$$\alpha_i = \frac{\ln(1 - F_i)}{\ln(x_0/x_{(i)})}, \quad (1)$$

где по-прежнему  $F_i = \frac{i}{n+1}$ . Задача сводится к оценке  $\hat{\alpha}$  как среднеарифметической оценке:

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \alpha_i. \quad (2)$$

При выборке объема  $n = 10$  из (1) и (2) получено  $\hat{\alpha} = 6.9$ , так что расчетное значение  $\hat{x}_{(10)} = 2.96$ , и точность аппроксимации повышается  $\Delta = \frac{3.12 - 2.96}{3.12} \cdot 100\% \approx 5\%$ .

Таким образом, распределение Парето удовлетворительно описывает экспериментальные данные в области их повышенных значений, что позволяет сделать предположение о возможности его применения в целях прогноза ВЗ и ЭВЗ.

## 2. Использование распределения Парето для прогноза ЭВЗ

Распределение Парето аналитически описывает функцию распределения, начиная с  $x_0$ . Остальная часть распределения представлена эмпирически. Поэтому требуется согласование его «хвостовой» части с основной. Для этого рассчитывается степень усечения  $F_{\text{ус.}} = \frac{n_e - n}{n_e + 1}$ , так что исходное распределение представляется в виде:

$$F = F_{\text{Парето}}(1 - F_{\text{ус.}}) + F_{\text{ус.}}. \quad (3)$$

**Пример 2.** Используя данные примера 1, где объем выборки  $n_e = 824$  при  $n = 20$  и  $n = 10$ , осуществить прогноз значений концентрации железа с риском ошибки  $r = 0.0001^2$ , т.е. на уровне  $F = 0.9999$ , что характеризует чрезвычайно редкое наступление данного события.

**Решение.** При  $n = 20$ :  $F_{\text{ус.}} = \frac{824 - 20}{825} = 0.975$ , и тогда прогноз концентрации железа на  $F = 0.9999$  осуществляется при  $F_{\text{Парето}} = \frac{0.9999 - 0.975}{1 - 0.975} = 0.996$ , так что по Парето  $\ln(1 - 0.996) = 6.32 \ln\left(\frac{1.8}{x_{\text{пр.}}}\right)$ . Следовательно, прогнозное значение предельной кон-

центрации  $x_{\text{пр.}(1)} = 4,31$  мг/дм<sup>3</sup>. Если же  $n = 10$ , прогноз на

$$F = 0.9999 \text{ рассчитываем при } F_{\text{ус.}} = \frac{824 - 10}{825} = 0.9866,$$

$F = \frac{0.9999 - 0.9866}{1 - 0.9866} = 0.9925$ , откуда при  $x_0 = 2.12$  мг/дм<sup>3</sup>,  $\hat{\alpha} = 6.9$  получим  $x_{\text{пр.}(2)} = 4,31$  мг/дм<sup>3</sup>. Как видно,  $x_{\text{пр.}(1)} = x_{\text{пр.}(2)}$ , что также иллюстрирует устойчивость полученных результатов.

В отличие от этого при использовании нормального закона распределения прогнозируемое значение при условиях примера 1 составляет:  $x_{\text{пр.}} = \hat{m} + u_{0.9999} \cdot \hat{\sigma} = 0.48 + 3.57 \cdot 0.469 = 2.15$  мг/дм<sup>3</sup>, где  $u_{0.9999} = 3.57$  – квантиль стандартного нормального распределения,  $\hat{m}$  и  $\hat{\sigma}$  – из примера 1. Как видно,  $x_{\text{пр.}}$ , найденное с использованием нормального закона, почти вдвое меньше найденного по Парето. Поэтому фильтры, отсекающие риск недопустимого загрязнения вод, рассчитанные на основе гипотезы о нормальном законе распределения вероятностей концентрации загрязняющих веществ, не обеспечивают необходимое качество фильтрации.

**Пример 3.** Избыточное содержание в воде многих взвешенных веществ промышленного происхождения оказывают сильное угнетающее действие на водные экосистемы. В то же время их концентрация характеризуется особенно повышенной нестабильностью из-за чего прогноз ЭВЗ для них наиболее затруднителен. Поэтому представляет интерес проверка возможностей предлагаемой методики прогноза ВЗ и ЭВЗ для взвешенных веществ. Требуется дать такой прогноз по данным для р. Исеть за 1986-2009 гг. на створах: 5.2 км выше г. Екатеринбурга, в черте города и 7.2 км ниже города.

**Решение.** Предварительно проведено исследование массивов данных на наличие аномальных выбросов по критерию Диксона: для одного выброса –  $\tau_1(n) = \frac{x_{(n_e)} - x_{(n_e-1)}}{x_{(n_e)} - x_{(1)}}$ , для двух выбросов –  $\tau_2(n) = \frac{x_{(n_e)} - x_{(n_e-2)}}{x_{(n_e)} - x_{(1)}}$  и т.д., где  $x_{(n_e)}, x_{(n_e-1)}, x_{(n_e-2)}, \dots, x_{(1)}$  – порядковые статистики каждой из трех заданных в примере выборок. При превышении  $\tau_1(n)$ ,  $\tau_2(n)$  установленных пороговых значений [14], соответствующие величины  $x_{(n_e)}, x_{(n_e-1)}$  исключались из рассмотрения как аномальные.

Возможность описания экспериментальных данных нормальным законом распределения проводилась по формулам:  $x_{(i)} = \hat{m} + u_{R(i)} \hat{\sigma}$ , где  $\hat{m} = \frac{1}{n_e} \sum_{i=1}^{n_e} x_i$ ;  $\hat{\sigma}^2 = \frac{1}{n_e - 1} \sum_{i=1}^{n_e} (x_i - \hat{m})^2$ ;  $R(i) = \frac{i}{n_e + 1}$ ,  $i = 1, 2, \dots, n_e$ , где  $x_i$  – измерения в каждой выборке объемом  $n_e$ ;  $u_{R(i)}$  – квантили стандартного нормального распределения.

2 Вероятности и риски даны со значениями четырех знаков после запятой, и не могут быть округлены без нарушения точности расчетов.

Таблица 2

Результаты статистического анализа наблюдений концентрации взвешенных веществ в р. Исеть. Размерность данных в строках 5, 7, 8, 9 – мг/дм<sup>3</sup>

Table 2

Results of the statistical analysis of the suspended solids observations in the Iset' River. The dimension of the data lines 5, 7, 8, 9 – in mg/dm<sup>3</sup>

№	Анализируемая характеристика	5.2 км выше Екатеринбурга	Створ в черте города	7.0 км ниже Екатеринбурга
1	Объем измерений	259	256	262
2	Число/значение аномальных выбросов	1/179	2/195, 382	2/407, 466
3	Доля случаев: меньше ПДК/ВЗ/ЭВЗ*	0.3/0.68/0.02	0.15/0.74/0.11	0.03/0.67/0.3
4	$\hat{m}/S$	5.8/12.04	17.1/29.85	18.87/40.1
5	Диапазон аппроксимации функции Парето	22÷67.8	92÷162.8	154÷314
6	Параметр $\alpha$	2.05	3.82	3.125
7	Максимум концентрации $\hat{x}_{(8)}/x_{(8)}$	67.7/67.8	168.2/162.8	311/314
8	Значения, обеспечивающие $r \leq 0.01$ для распределений: нормального/степенного/экспериментального	30/48.8/49	86.65/113/114	132.3/229.8/227.2
9	Прогноз концентрации, $r \leq 0.001$ . Распределения: нормальный/степенной	42/131	109.6/197	163.2/471

Примечание: \* – ПДК (здесь – предельно допустимая концентрация взвешенных веществ в воде водных объектов, используемых для рыбохозяйственных целей, второй категории) = 0.75 мг/дм<sup>3</sup>, ВЗ = 7.5 ÷ 37.5 мг/дм<sup>3</sup>, ЭВЗ > 37.5 мг/дм<sup>3</sup>.

Для прогноза наиболее высокого загрязнения воды выбирались подходящие объемы выборки таким образом, чтобы расчетное и экспериментальное максимальные значения при этом были достаточно близки. В данном примере такими оказались значения  $n = 8-9$ . Это – очень малый объем используемых данных, вследствие чего оценка параметра  $\alpha$  распределения проводилась относительно более точным методом наименьших квадратов по формуле (2), а согласование «хвостовой» и основной частей распределения – по (3).

Результаты расчета по данным примера 3 объединены в табл. 2

Как видно из суммирующей табл. 2, значения контролируемого показателя по Парето при  $r \leq 0.01$  близки к экспериментальным (строка 8). Это выгодно отличает степенной закон от экспоненциального и позволяет надеяться на достоверность первого также и при  $r > 0.01$ .

### 3. Оценка риска высокого загрязнения воды на основе теории порядковых статистик

Риск высокого загрязнения воды существует даже в случае относительно благополучных результатах периодических измерений. Водопользователю важно знать соответствующий риск, что требует оценки верхней границы доверительного интервала, покрывающего контролируемый показатель с приемлемой доверительной вероятностью. Если закон распределения неизвестен, в этом случае может быть использован метод, основанный на теории порядковых статистик.

Поскольку  $i$ -я порядковая статистика имеет приближенно нормальное распределение с математическим ожиданием  $x_{(i)}$  и среднеквадратическим отклонением  $\sqrt{nF_i(1-F_i)}$ , то для нахождения границ доверительного интервала необходимо изменить номер порядковой статистики на величину  $\Delta_i = \pm u_{\frac{1+\gamma}{2}} \sqrt{nF_i(1-F_i)}$  [13], где  $u_{\frac{1+\gamma}{2}}$  – квантиль стандартного нормального распределения,  $\gamma$  – доверительная вероятность. Найденные таким образом значения  $x_{H(i)} = x_{(i-\Delta_i)}$  и  $x_{B(i)} = x_{(i+\Delta_i)}$  принимаются, соответственно, за нижнюю и верхнюю доверительные границы для  $x_{(i)}$ .

Если оказывается, что  $x_{(n)} < x_{(i+\Delta_i)}$ , то на основе свойства симметричности нормального распределения в качестве верхней границы принимается значение  $x_{(i)} + [x_{(i)} - x_{(i-\Delta_i)}]$ . Если расстояние между соседними порядковыми статистиками большое, возможна интерполяция.

**Пример 4.** Рассчитать границы доверительного интервала непараметрическим методом для характеристик загрязнения воды по данным примера 3 на створе выше Екатеринбурга при заданном значении риска  $r = 0.01$ .

**Решение.** Здесь  $1 - r = \frac{i}{259} = 0.99$ , откуда  $i = 256.4$ . Значение статистики  $x_{(256)} = 41$ ;  $x_{(257)} = 57.1$ . Среднее значение, принятое в качестве оценки  $x_{(0.99)} = \frac{57.1 + 41}{2} \approx 49$ . Среднеквадратическое отклонение будет  $\sqrt{258 \cdot 0.99 \cdot 0.01} = 1.6$ ,  $\Delta_i = 1.64$ .

Таблица 3

Расчётные данные концентрации взвешенных веществ в р. Исеть, возникающие с риском не выше 0.01

Table 3

Estimated data of the suspended matter concentration in the Iset' River arising with a risk of not more than 0.01

Измерения	Порядковые статистики			Аппроксимация распределением Парето					
	Оценка	$x_n$	$x_b$	$\alpha$			Оценка	$x_n$	$x_b$
				Оценка	$\alpha_n$	$\alpha_b$			
5.2 км выше по течению	49	38	67.8	2.05	< 2	2.55	48.8	38	58
В черте города	133	112	162.8	3.82	< 2	6.5	129	115	143
7.0 км ниже по течению	224.4	186	314	2.42	< 2	2.8	229.8	217	242.6

$1.6 = \pm 2.6 \approx 3$  при  $\gamma = 0.9$ , откуда, при соответствующем округлении из эмпирической функции распределения получено:  $x_{n(0.99)} = 38$ ,  $x_{b(0.99)} = 67.8$ .

Как видно, даже очень высокие загрязнения воды, отмеченные в табл. 2 в течение учтенных 23 лет могут происходить чаще, чем это следует из данных гидромониторинга.

В ответственных случаях целесообразно уточнение границ доверительного интервала, для чего в данной задаче воспользуемся аппроксимирующим распределением Парето, и оценим его параметр  $\alpha$  методом наименьших квадратов. Такая оценка величины  $\hat{\alpha}$  является среднеарифметической, что на основе центральной предельной теоремы позволяет применить нормальный закон распределения вероятностей с математическим ожиданием  $\alpha$  (истинное значение) и оценкой дисперсии  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (\alpha_i - \hat{\alpha})^2$ . Границы доверительного интервала  $[\alpha_b; \alpha_n]$  в этом случае определяются как  $\hat{\alpha} \pm u_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}$ , а соответствующие им  $x_{n(i)}$  и  $x_{b(i)}$  рассчитываются с использованием выражения (1) при подстановке в него значений  $\alpha_n$  и  $\alpha_b$ .

Заметим, что поскольку дисперсия по Парето равна  $S^2_{\text{Парето}} = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} x_0^2$ , это распределение оказывается определено при  $\alpha > 2$ , поэтому если оценка  $\hat{\alpha} < 2$ , рассчитывается только  $x_{n(i)}$ , а далее, вследствие симметричности нормального распределения, значение  $x_{b(i)}$  определяется по алгоритму, указанному выше.

**Пример 5.** Решить пример 4, используя параметрический метод.

**Решение.** При условиях примера 4 оценка параметра распределения Парето  $\hat{\alpha} = 0.05$ ;

$$S = 0.82, \quad \alpha_n = 2.05 - \frac{1.64 \cdot 0.82}{3} = 1.6 < 2;$$

$$\alpha_b = 2.05 + \frac{1.64 \cdot 0.82}{3} = 2.5; \quad x_{n(0.99)} = 38.3;$$

$$x_{b(0.99)} = (48 - 38.3) + 48 = 57.7.$$

Для малого объема экспериментальных данных распределение  $\hat{\alpha}$  будет уже не нормальным, а

$t$ -распределением Стьюдента, т.е. доверительный

интервал определяется как  $\hat{\alpha} \pm t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}$ , где  $t_{\frac{1+\gamma}{2}}$  - квантиль распределения Стьюдента. В результате получено:  $\alpha_n = 2.05 - \frac{1.8 \cdot 0.82}{3} = 1.56 < 2$ ;  $x_{n(0.99)} = 38$ ;

$x_{b(0.99)} = 58$ . Таким образом, значения нижней доверительной границы, полученные непараметрическим и параметрическим методами, практически совпадают, однако, интересующая нас верхняя граница при использовании распределения Парето несколько снижена, что следует из табл. 3.

Как видно из табл. 3, строки 3, 5, снижение верхней границы при использовании распределения Парето тем заметнее, чем ниже по течению реки выбран створ т.е. в этом направлении нарастает нестабильность контролируемых показателей.

Сужение доверительного интервала при переходе от непараметрического метода оценивания к параметрическому позволяет снизить меры предосторожности, необходимые на случай повышенного загрязнения вод.

## Заключение

Различие результатов, получаемых при использовании для целей анализа, контроля и прогноза высокого загрязнения вод нормального или степенного распределений вероятностей носит принципиальный характер. Если статистика случайной величины описывается нормальным законом, то в более чем 99.7 % событий величина отклоняется от среднего значения менее чем на  $3\sigma$  ( $\sigma$  – стандартное отклонение), а, например, за границу  $5\sigma$  – реже, чем в одном случае из миллиона. Поэтому очень «крупными» событиями, когда случайная величина достигает значений, много больших среднего, можно пренебречь, считая их практически невозможными. Если использовать для стандартной обработки данных мониторинга распределение из семейства экспоненциальных, например, как рекомендовано в [15], то катастрофические события (наводнения, загрязнения) будут всегда неожиданными. Наводнения исключительной силы последних лет убедительно показали,

что рассчитывать узлы водоподготовки, защитные дамбы, плотины и другие гидротехнические сооружения необходимо на основании иных вероятностных закономерностей.

Аналитическое описание «хвостовых» частей распределения вероятностей исследуемых показателей качества воды распределением Парето позволяет по выборкам ограниченного объема определять соответствующие показатели в ненаблюдаемых частях распределений – давать прогнозные оценки. Область рационального применения данного подхода – оценивание малых значений рисков по небольшому количеству редких событий.

Использование параметрического подхода в виде сочетания усеченной эмпирической формулы распределения и распределения Парето для аналитического описания хвостовой части позволяет разработать простые прикладные методики расчета, не требующие сложного программного обеспечения. А вариации объема этой хвостовой части из условия обеспечения максимального совпадения расчетных и экспериментальных данных позволяют предположить универсальный характер данного подхода, что подтверждается широкой его апробацией в задачах других предметных областей, связанных с необходимостью оценки сверхмалых рисков, характеризующих безопасность процессов различной природы [12].

## ЛИТЕРАТУРА

1. РД 52.24.643-2002. Методические указания. Метод комплексной оценки степени загрязненности поверхностных вод по гидрохимическим показателям. С.-Пб.: Гидрометеиздат, 2002. 55 с.
2. Р 52.24.627-2007. Рекомендации. Усовершенствованные методы прогностических расчетов распространения по речной сети зон высоко-загрязненных вод с учетом форм миграции наиболее опасных загрязняющих веществ. Ростов н/Д, 2007. 172 с.
3. Р 52.24.756-2011. Критерии оценки опасности токсического загрязнения поверхностных вод суши при чрезвычайных ситуациях (в случаях загрязнения). Ростов н/Д, 2011. 43 с.
4. Р 52.24.734-2010. Организация и проведение наблюдений за состоянием и изменением качества поверхностных вод в чрезвычайных ситуациях. Ростов н/Д, 2011. 66 с.
5. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998. 1000 с.
6. Авербух А.И., Розенталь О.М. Унификация эмпирических распределений показателей качества воды на основе нормального и степенного законов. // Водное хозяйство России: проблемы, технологии, управление. 2014. № 2. С. 60-68.
7. Александровская Л.Н., Розенталь О.М. Сравнение методов оценивания риска ошибочного контроля качества природных вод. // Аналитика и контроль. 2016. Т. 20, №1. С. 53-61.
8. Беврани Х., Аничкин К. Оценка параметров распределений с тяжелыми хвостами с помощью эмпирического распределения. // «Математика. Компьютер. Образование». Сб. трудов XII международной конференции. Под общ. редакцией Г.Ю. Ризниченко. Ижевск: Научно-издатель-

- ский центр «Регулярная и хаотическая динамика», 2005. Т. 2. С. 493-501.
9. Taqqu M., Teverovsky V., Willinger W. Estimators for long-range dependence: an empirical study. // *Fractals*. 1995. V. 3, № 4. P. 785-798.
10. Крицкий С.Н., Менкель М.Ф. Гидрологические основы управления речным стоком. М.: Наука, 1981. 257 с.
11. Сикан А.В. Практические приемы оценки параметров распределения Вейбулла при выполнении гидрологических расчетов. // Ученые записки Российского государственного гидрометеорологического университета. 2011. №19. С. 37-46.
12. Метод прогнозирования ненаблюдаемых «хвостов» функций распределения в задачах оценки соответствия требований к безопасности систем автоматической посадки самолетов нормам летной годности / Л.Н. Александровская [и др.] // Труды МИЭА. Навигация и управление летательными аппаратами. 2012. №5. С. 38-54.
13. Дейвид Г. Порядковые статистики: пер. с англ. М.: Наука. 1979. 336 с.
14. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
15. СНиП 2 01.14-83. Определение расчетных гидрологических характеристик. М.: Стройиздат, 1985.

## REFERENCES

1. RD 52.24.643-2002. *Metodicheskie ukazaniia. Metod kompleksnoi otsenki stepeni zagriznennosti poverkhnostnykh vod po gidrokhimicheskim pokazateliyam*. [Directive Document RD 52.24.643-2002. The method of complex assessment of surface water contamination degree by hydrochemical indicators]. Saint Petersburg, Gidrometeoizdat Publ., 2002. 55 p. (in Russian).
2. R 52.24.627-2007. *Rekomendatsii. Usovershenstvovaniye metody prognosticheskikh raschetov rasprostraneniia po rechnoi seti zon vysoko-zagriznennykh vod s uchetom form migratsii naibolee opasnykh zagrizniaiushchikh veshchestv*. [Recommendations R 52.24.627-2007. Advanced methods of predictive calculation of highly contaminated water propagation to the river network areas with regard to the forms of migration of the most dangerous pollutants]. Rostov-on-Don, 2008. 172 c. (in Russian).
3. R 52.24.756-2011. *Kriterii otsenki opasnosti toksicheskogo zagrizneniia poverkhnostnykh vod sushi pri chrezvychainykh situatsiiakh (v sluchaiakh zagrizneniia)* [Recommendations R 52.24.756-2011. Criteria for assessing the risk of toxic contamination of surface waters in emergency situations (in case of pollution)]. Rostov-on-Don, 2011. 43 p. (in Russian).
4. R 52.24.734-2010. *Organizatsiia i provedenie nabludeniia za sostoianiem i izmeneniem kachestva poverkhnostnykh vod v chrezvychainykh situatsiiakh* [Recommendations R 52.24.734-2010. Organizing and performing study of the status and changes in the surface water quality in emergency situations]. Rostov-on-Don 2011. 66 p. (in Russian).
5. Aivazian S.A., Mkhitarian V.S. *Prikladnaia statistika i osnovy ekonometriki* [Applied Statistics. Basics of econometrics]. Moscow, UNITY, 1998, 1000 p. (in Russian).
6. Averbukh A.I., Rosental' O.M. [Unification of the empirical distributions of water quality parameters based on the normal and power laws]. *Vodnoe hoziaistvo Rossii: problemy, tehnologii, upravlenie* [Water sector of Russia: problems, technologies, management], 2014, no. 2. pp. 60-68 (in Russian).
7. Aleksandrovskaia L.N., Rosental' O.M. [Comparison of the risk estimation methods for erroneous monitoring of natural waters quality]. *Analitika i kontrol'* [Analysis and Con-

- trol], 2016, vol. 20, no. 1, pp. 53-61. DOI: 10.15826/analitika.2015.20.1.001 (in Russian).
8. Bevrani H., Anichkin K. [Parameter estimation for the heavy tailed distributions with the empirical distribution]. «*Matematika. Komp'yuter. Obrazovanie*». Sb. trudov XII mezhdunarodnoi konferentsii [Proc. 7<sup>th</sup> International Conference «Mathematic. Computer. Education»]. Izhevsk, 2005. vol. 2, pp. 493-501. (in Russian).
9. Taqqu, M., Teverovsky, V., and Willinger, W. Estimators for long-range dependence: an empirical study. *Fractals*, vol. 3, no. 4, pp. 785–798. DOI 10.1142/S0218348X95000692.
10. Krickij S.N., Menkel' M.F. [Hydrological foundations of river flow management]. Moscow, Nauka Publ., 1981. 257 p. (in Russian).
11. Sikan A.V. [Practical methods of assessment of the Weibull distribution parameters in the performance of hydrological calculations]. *Uchenye zapiski Rossiiskogo gosudarstvennogo gidrometeorologicheskogo universiteta* [Scientific notes of Russian State Hydrometeorological University]. 2011, no. 19. pp. 37-46. (in Russian).
12. Aleksandrovskaya L. N., Mazur, V. N., Khlghatyan, S. V., Ardalionovna A. E. [Method of prediction non-visible function's 'tails' distribution in the problem of accordance evaluation of the terms to automatic landing systems safety with the norms of airworthiness]. *Trudy MIEA. Navigatsiia i upravlenie letatel'nyimi apparatami* [Works of the MIEA. Navigation and control of flight vehicles]. 2012, no. 5. pp. 38-54. (in Russian).
13. David, H.A. *Order statistics*. Wiley, New York, 1970, 272 p. (Russ ed: David H. Poriadkovye statistiki. Moscow, Nauka Publ., 1979, 336 p. (in Russian)).
14. Bol'shev L.N., Smirnov N.V. *Tablitsy matematicheskoi statistiki* [Tables of mathematical statistics]. Moscow, Nauka Publ., 1983, 416 p. (in Russian).
15. SNIP 2 01.14-83. *Opredelenie raschetnykh gidrologicheskikh kharakteristik* [Construction Norms & Regulations 2 01.14-83. Determination of design hydrological characteristics]. Moscow, Strojizdat Publ., 1985.